SAMMate

Computational Biology Group
Department of Computer Science
University of New Orleans

May 2011

[MANUAL]

SAMMate allows biomedical researchers to quickly process Fasta/Fastq, SAM/BAM files. This software is constantly updated and will greatly facilitate the downstream analysis of NGS data in the Fasta/Fastq, SAM/BAM format. Both the source code and the GUI executable are freely available at http://SAMMate.sourceforge.net.

SAMMate Manual

Copyright © 2009-2011 Computational Biology Group@UNO Computer Science. All rights reserved.

Abstract

Welcome to the *SAMMate* Manual. Here you will find information on how to install and configure the application. It is a step-by-step, task-oriented guide for configuring *SAMMate* on your system.

License

This document is maintained by the Computational Biology Group at UNO Computer Science and is freely available under the GNU General Public License.

SAMMate is free software; you can redistribute it and/or modify it under the terms of version 2 of the GNU General Public License as published by the Free Software Foundation.

SAMMate is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

A copy of version 2 of the GNU General Public License is appended in the installation package. For more information, see <http://www.gnu.org/licenses/>.

Table of Contents

Chapter 1 Introduction 5
Chapter 2 Installation 6
2.1. Preparing to Install 7
2.1.1. System Requirements 7
2.1.2. Running Environment 7
2.2 Installing SAMMate 8
Chapter 3 Usages 9
3.1. File Management 10
3.1.1. Working Directory 10
3.1.2. Work Space 11
3.2. Run Options Configuration 13
3.2.1. Run Options 13
3.2.2 SAM/BAM File Operations 14
3.2.3 Bowtie Options Configuration 15
3.2.4 R Options Configuration 16
3.3. SAM/BAM Format Conversion 18
3.3.1 Format Conversion 18
3.4. SAM/BAM File Sorting 19
3.4.1 Sorting a SAM/BAM File 19
3.5. Gene Expression Abundance Score Calculation 20
3.5.1 Processing a Genome Annotation File and RNA-seq Data Files 20
3.6. File Format Customization 21
3.6.1 Customizing a genome annotation file 21
3.6.2 Customizing the signal map intervals 21
3.6.3 Customizing the chromosome names 22
3.7. Memory Configuration on MAC OS 24
3.7.1 Allocating more memory on the MAC OS 24
Chapter 4 Case Studies 25
4.1. Estimation of Transcripts Abundance 26
4.1.1 Example 26
4.2. Calculating Genomic Feature Abundance Scores 28

4.2.1 Example 28
4.3. Generating a Signal Map for Peak Detection 30
4.3.1 Example 30
4.4. Generating Wiggle File for Visualization 32
4.4.1 Example 32
4.5. Generating Alignment Report 34
4.5.1 Example 34

Chapter 1 Introduction

Welcome to the *SAMMate* Manual.

Here you can find information on how to install and configure SAMMate. It is a step-by-step, task-oriented guide for configuring *SAMMate* on your system.

This manual assumes you have a basic understanding of your operating system. Some installation details are covered in Chapter 2: Installation. If you need detailed instructions on using *SAMMate*, please refer to Chapter 3: Usages and Chapters 4: Use Case Studies.

An HTML version of the manual is available online at SAMMate's homepage:

http://SAMMate.sourceforge.net/

Chapter 2 Installation

This chapter provides a quick overview on installing *SAMMate*.

2.1. Preparing to Install

This section explains SAMMate's requirements.

2.1.1. System Requirements

Recommended Memory: 2GB RAM
Minimum Memory: 1GB RAM
OS: Windows 7, Vista, XP, Mac OS X, Linux OS

2.1.2. Running Environment

A recent version of the Java Runtime Environment (JRE) is needed prior to using SAMMate.

For Windows this involves a quick download from the following Oracle site:
http://www.java.com/en/download/manual.jsp

Java is standard on Mac OS X. For Mac OS 9, you may need to get the Mac OS Runtime for Java (MRJ) from the following Apple site:
http://docs.info.apple.com/article.html?artnum=75097

2.2 Installing SAMMate

- Download the zip file *SAMMate2.6.1.zip* that matches your OS to your local hard drive.
- Decompress it, and open the software folder *SAMMate 2.6.1*.
- Double click the executable file *SAMMate.exe* (Windows), *SAMMate.app* (Mac) or *SAMMate.sh* (Linux).

Chapter 3 Usages

This chapter provides a detailed guide of using *SAMMate*. Some typical usage scenarios are listed below:

- Working directory management.
- Adding annotation files and data files to the workspace.
- Removing annotation files and data files from the workspace.
- Configuring run options
- Configuring Bowtie options
- Configuring R options
- Converting from SAM format to BAM format and vice-versa.
- Sorting SAM/BAM files.
- Customization of Genome Annotation File
- Customization of Signal Map Intervals
- Customization of Chromosome Names

3.1. File Management

You can change the working directory and add/remove the annotation file or data files to/from the *Work Space* in SAMMate.

3.1.1. Working Directory

Users can change the working directory for managing the annotation file and data files or for outputting the resulting files.

To display the directory dialogue

- Select File > Open from the menu, or click the  toolbar button.

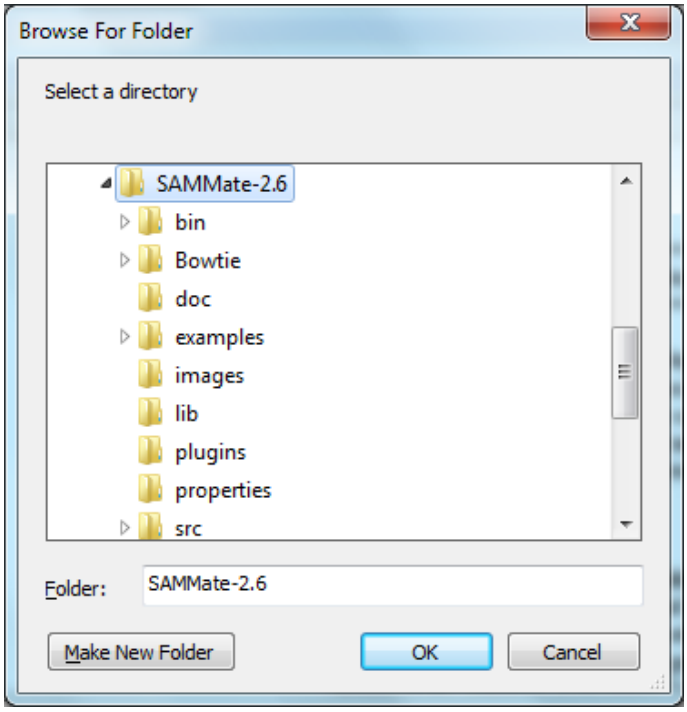


Figure 3.1.1-1 Open directory dialogue

- From the directory tree, select the desired folder, and press the OK button.
- The directory and the files under the selected folder are displayed in the *File Browser Window*.

3.1.2. Work Space

Work Space allows users to add/remove the annotation file and data files.

To add files to the Work Space

- In the *File Browser Window*, right click on the desired annotation or data file, and then left click *Add to Work Space*. Alternatively, double left click on the desired file in the *File Browser Window*. Either method will add the selected file to the *Work Space*.

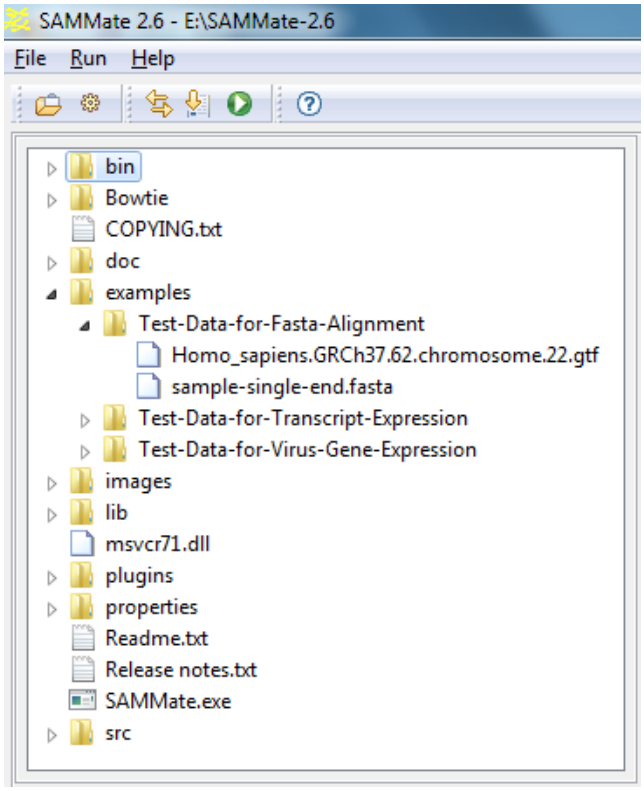


Figure 3.1.2-1 File Browser window

To remove files from the Work Space

- In the *Work Space*, right click on the desired annotation or data file, and then left click *Delete Selection*. Alternatively, double left click on the desired file to remove it from the *Work Space*.

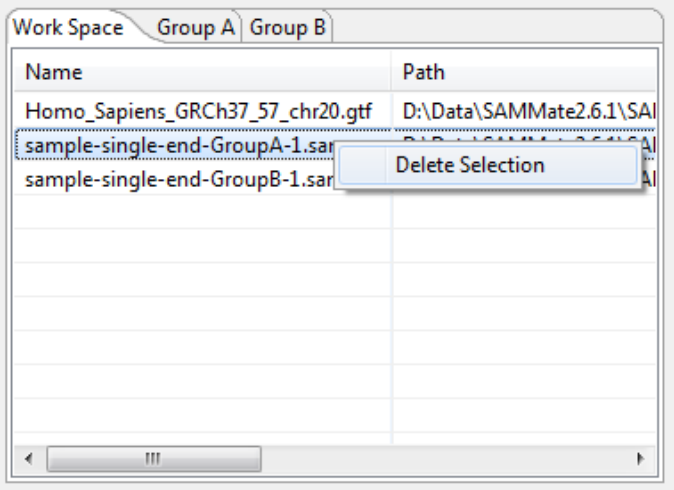


Figure 3.1.2-2 Work Space

3.2. Run Options Configuration

You can configure the run options to enable/disable the key features on the *Options* dialogue, or customize the output file names in SAMMate.

3.2.1. Run Options

SAMMate allows users to estimate transcripts abundance using different methods, to enable/disable the key features and customize output file name.

To estimate transcript abundance

- Select File > Options from the menu, or click the toolbar button.
- A popup dialogue appears with different methods of estimating transcripts abundance already loaded on *Run Options* page. Three methods are listed: one-step iQuant, iterative iQuant and RAEM algorithm. See Figure 3.2.1-1.

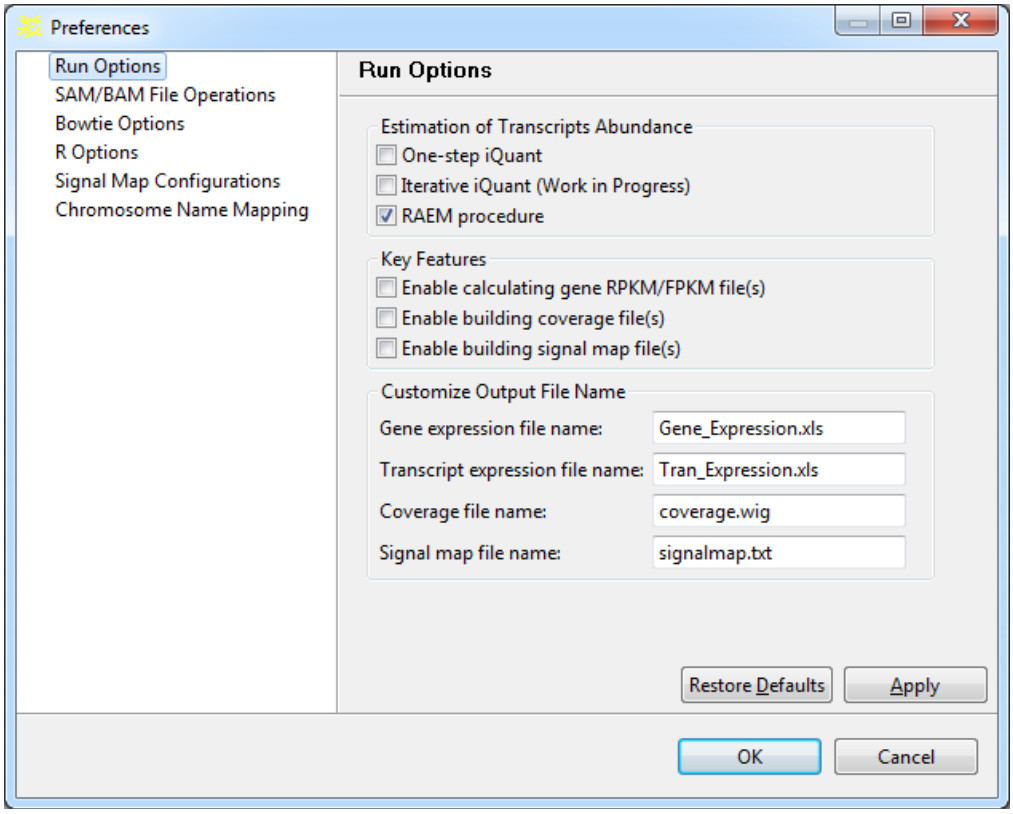


Figure 3.2.1-1 Options dialogue – Run Options

To enable/disable key features

- Select File > Options from the menu, or click the toolbar button.

- A popup dialogue appears with key features already loaded on *Run Options* page. SAMMate allows user to enable/disable calculating gene RPKM/FPKM file, enable/disable building coverage file and enable/disable building signal map file. See Figure 3.2.1-1.

To customize output file name

- Select File > Options from the menu, or click the toolbar button.
- A popup dialogue appears with customization of output file name already loaded on *Run Options* page. SAMMate allows user to customize the output file name. See Figure 3.2.1-1.

3.2.2 SAM/BAM File Operations

SAMMate allows users to set three sorting orders when user sorts SAM/BAM file: unsorted, query name and coordinate.

To set sorting order

- Select File > Options from the menu, or click the toolbar button.
- Clicking *SAM/BAM File Operations*, users can sort SAM/BAM file by the reference coordinates, by query names or unsorted by default.

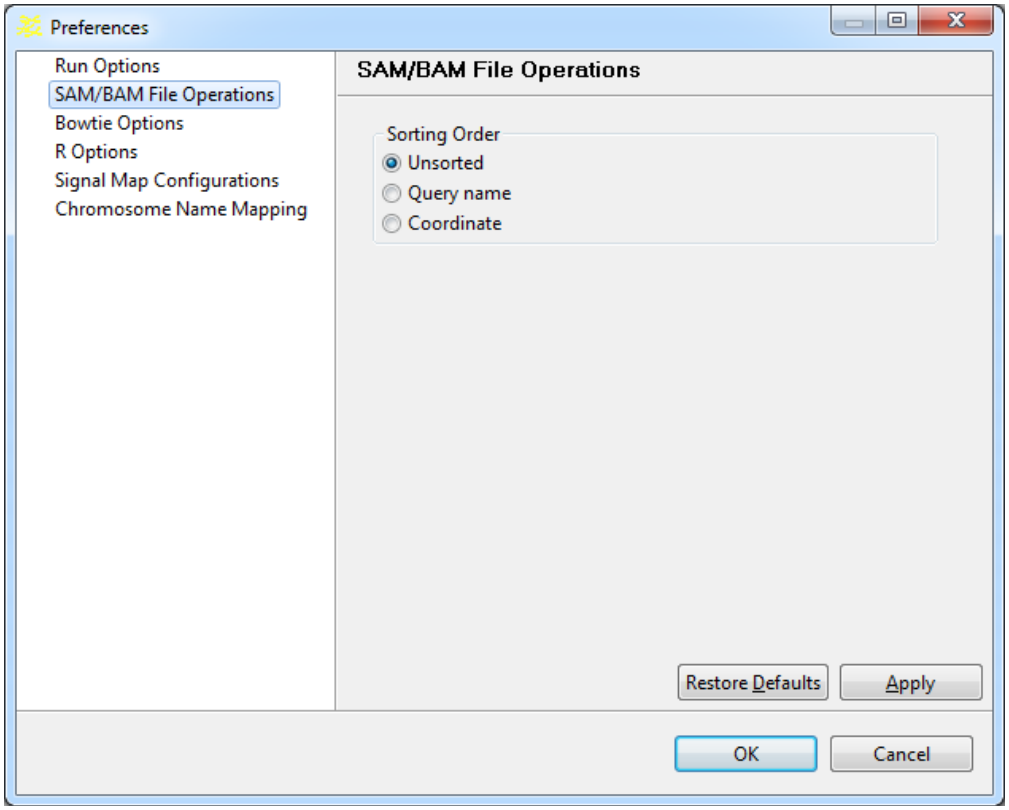


Figure 3.2.2-1 Options dialogue

3.2.3 Bowtie Options Configuration

SAMMate allows users to configure Bowtie aligner to process Fasta or Fastq sequence file.

To configure Bowtie options

- Select File > Options from the menu, or click the toolbar button.
- Clicking *Bowtie Options*, users can configure path of Bowtie installation, path of index files, and main arguments.

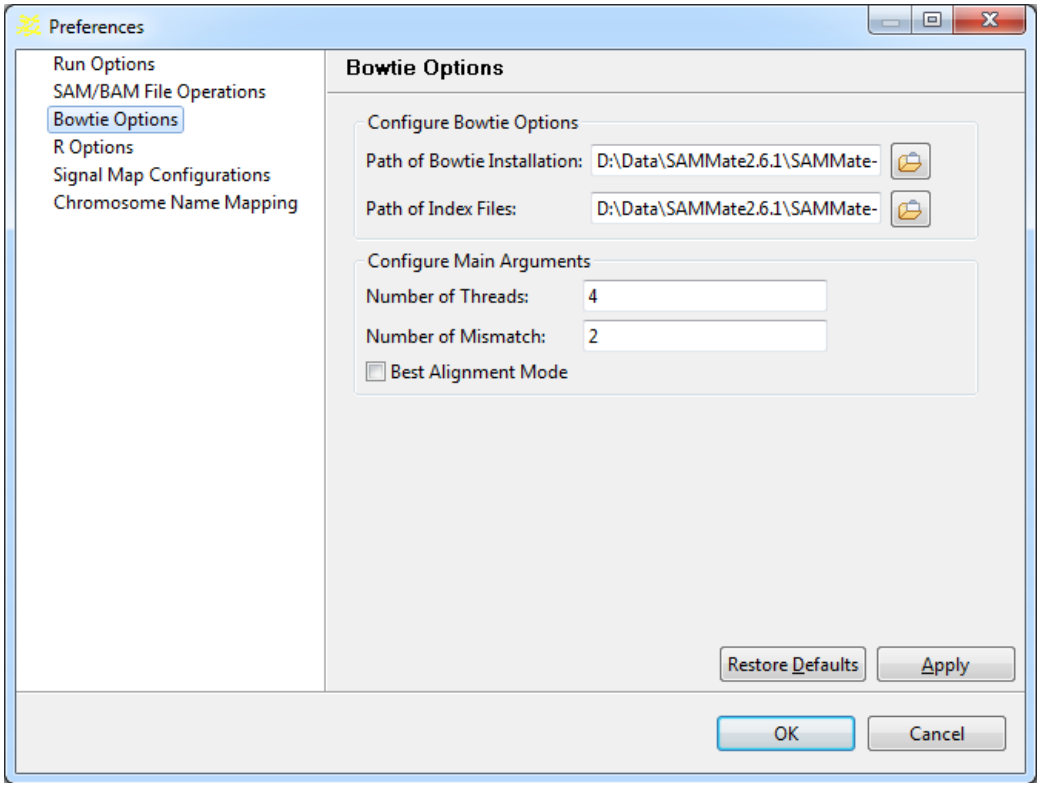


Figure 3.2.3-1 Options dialogue – Bowtie Options

- In the *File Browser Window*, right click on the desired Fasta/Fastq file, and then left click *Add to Work Space*.
- In the *Work Space*, you can right click on the Fasta/Fastq file, and then left click *Delete Selection* to remove selected file from *Work Space*.
- Select Run > Run from the menu, or click the toolbar button to align the Fasta/Fastq files present in the table.

For large sequence file, the Mac and Linux version of SAMMate are recommended.

3.2.4 R Options Configuration

SAMMate allows users to configure edgeR package to detect differentially expressed genes and isoforms.

To configure R options

- Select File > Options from the menu, or click the toolbar button.
- Clicking *R Options*, users can enable using edgeR, customize the output file names, and configure the path of R executable file (for example, Rscript.exe on Windows platform and Rscript on Mac/Linux platform).

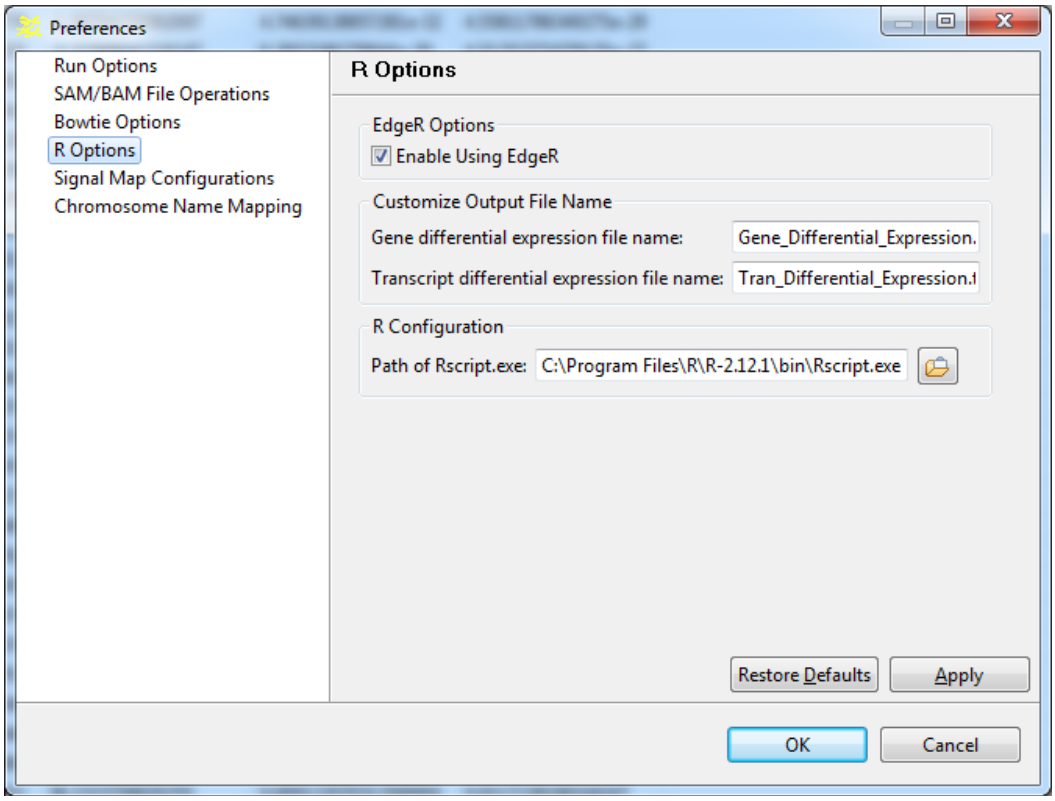


Figure 3.2.4-1 Options dialogue – R Options

- In the *File Browser Window*, right click on the desired sequence files, and then left click *Add to Work Space*.
- In the *Work Space*, you can right click on the selected files, and then left click *Delete Selection* to remove selected file from *Work Space*.
- In the group A and group B tabs, you can assign the selected files into different groups.
- Select Run > Run from the menu, or click the toolbar button.
- After the process is finished, the gene differential expression and transcript differential expression scores are displayed in the *Navigator window*. Two text files

compatible gene differential expression matrix and transcripts differential expression matrix are also generated in the temporary folder.

3.3. SAM/BAM Format Conversion

SAM is a TAB-delimited text format that is easy to understand, parse, generate and check for errors. However, for extremely large file sizes, SAM is a bit slow to parse. Binary SAM, i.e. BAM, is often used in many production pipelines for intensive data processing.

3.3.1 Format Conversion

SAMMate allows users to convert a file from the SAM format to the Binary SAM (BAM) format and *vice versa*.

To convert SAM/BAM format

- In the *File Browser Window*, right click on the desired data file, and then left click *Add to Work Space*.
- In the *Work Space*, you can right click on the SAM/BAM file, and then left click *Delete Selection* to remove selected file from *Work Space*.
- Select Run > Convert between SAM and BAM Format from the menu, or click the  toolbar button to convert the SAM/BAM files present in the table.

3.4. SAM/BAM File Sorting

A SAM/BAM file can be sorted by the reference coordinates, by query names, or unsorted. Sorting the SAM/BAM file is a crucial step for data processing on a stream and for indexing.

3.4.1 Sorting a SAM/BAM File

SAMMate allows users to sort data files in SAM format or Binary SAM (BAM) format.

To sort a SAM/BAM file

- In the *File Browser Window*, right click on the desired SAM/BAM file, and then left click *Add to Work Space*.
- In the *Work Space*, you can right click on the selected file, and then left click *Delete Selection* to remove selected file from *Work Space*.
- Select File > Options from the menu, or click the  toolbar button.
- You can check the desired order of sorting (See 3.2.2 SAM/BAM File Operations).
- Select Run > Sort SAM/BAM files from the menu, or click the  toolbar button to sort the SAM/BAM files present in the table.

3.5. Gene Expression Abundance Score Calculation

Using the standard reference genome annotation files, *SAMMate* allows users to accurately calculate the gene expression abundance scores for all annotated genes using RNA-seq data.

3.5.1 Processing a Genome Annotation File and RNA-seq Data Files

SAMMate is able to use short reads originating from both exons and exon-exon junctions to accurately calculate gene expression scores, build coverage depth and generate the signal map for peak detection.

To calculate the gene expression abundance score

- In the *File Browser Window*, right click on the desired data file, and then left click *Add to Work Space*.
- In the *Work Space*, you can right click on the SAM/BAM file, and then left click *Delete Selection* to remove selected file from *Work Space*.
- Select File > Options from the menu, or click the  toolbar button.
- You can check the desired features to enable calculating gene RPKM/FPKM file(s) (See 3.2.1 Run Options).
- Select Run > Run from the menu, or click the  toolbar button.



The selected files ***must*** include only one genome annotation file and at least one SAM/BAM file. The BED file is optional.

3.6. File Format Customization

3.6.1 Customizing a genome annotation file

SAMMate also calculates the abundance scores for customized genomic intervals by customizing a genome annotation file.

- Go to the *SAMMate 2.6.1 > example > Test-Data-for-Transcript-Expression* folder and open one genome annotation file as the reference.
- Follow the format in the selected annotation file to customize a genome annotation file by adding the gene name, chromosome name, exon count, exon start position and end position, etc.
- Add the customized genome annotation file and RNA-seq data files into the *Work Space* to calculate the gene expression abundance score, coverage depth, signal map and so on.

#geneName	name	chrom	strand	txStart	txEnd	cdsStart	cdsEnd	exonCount	exonStarts	exonEnds
FAM138F	NR_026820	chr1	-	34611	36081	36081	36081	3	34611,35276,35720,	35174,35481,36081,
FAM138A	NR_026818	chr1	-	34611	36081	36081	36081	3	34611,35276,35720,	35174,35481,36081,
FAM138C	NR_026822	chr1	-	34611	36081	36081	36081	3	34611,35276,35720,	35174,35481,36081,
OR4F5	NM_001005484	chr1	+	69090	70008	69090	70008	1	69090,	70008,
LOC100132287	NR_028322	chr1	+	323891	328580	328580	328580	3	323891,324287,324438,	324060,324345,328580,
LOC100132062	NR_028325	chr1	+	323891	328580	328580	328580	3	323891,324287,324438,	324060,324345,328580,
OR4F29	NM_001005221	chr1	+	367658	368595	367658	368595	1	367658,	368595,
OR4F3	NM_001005224	chr1	+	367658	368595	367658	368595	1	367658,	368595,
OR4F16	NM_001005277	chr1	+	367658	368595	367658	368595	1	367658,	368595,
OR4F29	NM_001005221	chr1	-	621097	622034	621097	622034	1	621097,	622034,
OR4F3	NM_001005224	chr1	-	621097	622034	621097	622034	1	621097,	622034,
OR4F16	NM_001005277	chr1	-	621097	622034	621097	622034	1	621097,	622034,

Figure 3.6.1-1 Customizing genome annotation file

3.6.2 Customizing the signal map intervals

SAMMate allows user to customize the signal map intervals to generate the base-wise signal map information that fall between the customized the intervals for peak detection.

- Select File > Options from the menu, or click the  toolbar button.
- Clicking *Signal Map Configurations*, users can customize the signal map intervals by adding the chromosome name, start position and end position. See Figure 3.6.2-1.
- In the *Signal Map* table, right click on the desired interval, and then left click *Delete Selection*. Alternatively, double left click on the desired file to remove it from the table.

- You can check the desired features to *enable building signal map file* (See 3.2.1 Run Options).
- In the *File Browser Window*, right click on the desired data file, and then left click *Add to Work Space*.
- In the *Work Space*, you can right click on the SAM/BAM file, and then left click *Delete Selection* to remove selected file from *Work Space*.
- Select Run > Run from the menu, or click the toolbar button to build the signal map file.

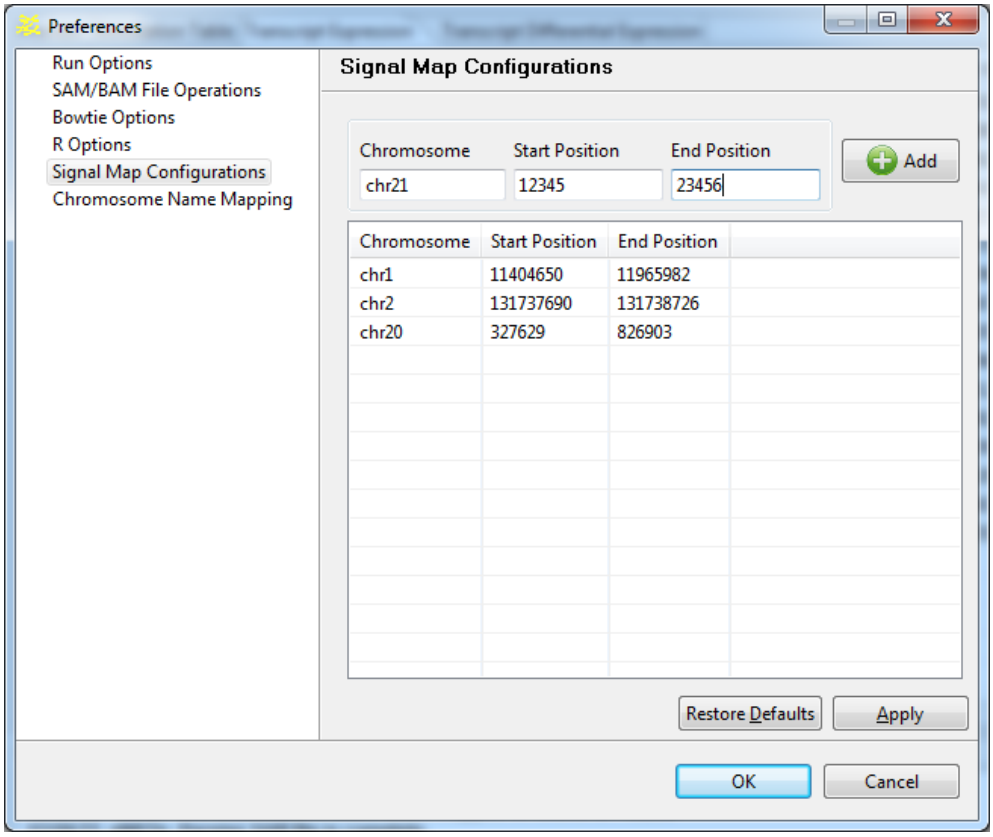


Figure 3.6.2-1 Options dialogue - Signal map configurations

3.6.3 Customizing the chromosome names

Between the genome annotation file and the RNA-seq data file, the chromosome names are often mismatched due to different databases and/or aligners. To remedy this situation, *SAMMate* allows user to customize the relationship map between different chromosome names allowing the system to automatically map the customized chromosome names during calculations.

- Select File > Options from the menu, or click the toolbar button.
- Clicking *Chromosome Name Mapping*, users can define the mapping relationship of chromosome names between different versions.
  For example, by adding add the line:

  gi|89161185|ref|NC_000001.9|NC_000001 chr1

  SAMMate will automatically replace the string "gi|89161185|ref|NC_000001.9|NC_000001" with "chr1" in the output files.
- In the *Chromosome Name Mapping* table, right click on the desired names, and then left click *Delete Selection*. Alternatively, double left click on the desired file to remove it from the table.

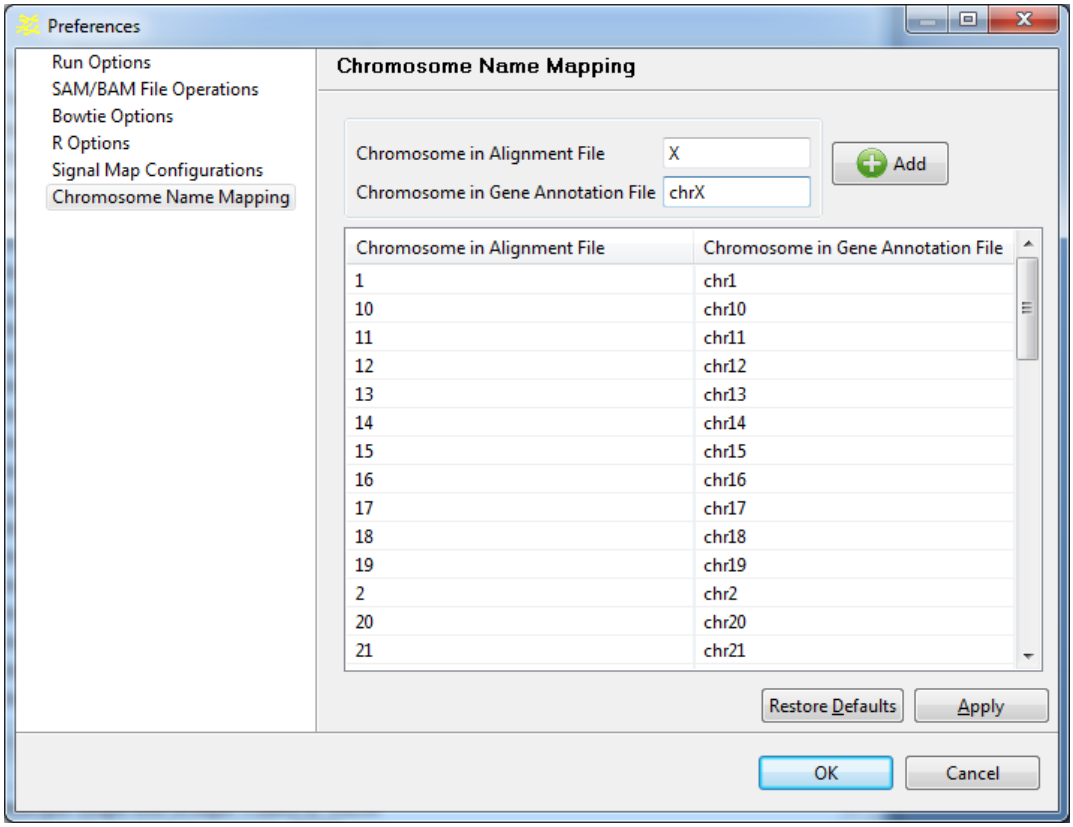


Figure 3.6.3-1 Options dialogue - Chromosome name mapping

3.7. Memory Configuration on MAC OS

3.7.1 Allocating more memory on the MAC OS

SAMMate also allows users to increase the amount of memory allocated to improve its performance on the Mac OS platform. Users can easily increase the amount of memory to use by modifying the launching file (*SAMMate.app*).

- Go to the *SAMMate 2.6.1* folder, and right click on the launching file *SAMMate.app*.
- Select "*Show Package Contents*".
- Go to the folder *Contents*.
- Open the file "*Info.plist*" by double clicking on it.
- In the line Java -> VMOptions, change the parameter from "-Xmx1024M" to "-Xmx2048M".
- Save the changes, and launch *SAMMate.app* again.

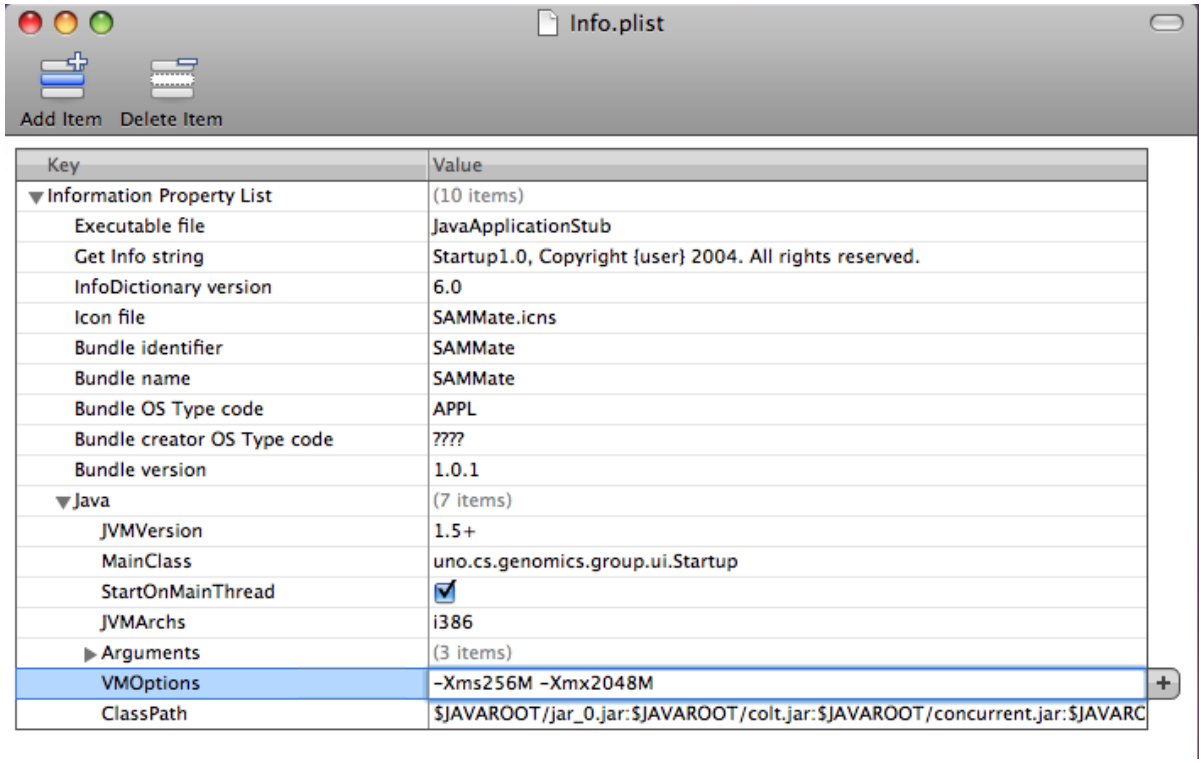


Figure 3.7.1-1 Info.plist file

On a 32-bit machine, the allocated memory cannot exceed 2048MB (or 2GB).

Chapter 4 Case Studies

This chapter provides some detailed examples for user case studies:

- Estimating transcripts abundance scores
- Calculating genomic feature abundance scores
- Generating a signal map for peak detection
- Generating coverage wiggle files for visualization
- Generating an alignment report

The genome annotation file and simulation RNA-seq data files used in these examples can be downloaded from

http://SAMMate.sourceforge.net/download.html

4.1. Estimation of Transcripts Abundance

SAMMate is compatible with both single-end and paired-end short reads mapped to exons (e.g. available in SAM/BAM format) to accurately estimate transcript abundance using three different methods: one-step iQuant, iterative iQuant and RAEM algorithm. *SAMMate* can also process RNA-seq data file in BED format.

4.1.1 Example

Download the genome annotation file and the simulation RNA-seq data to a local hard drive from http://SAMMate.sourceforge.net/download.html. Unzip to a temporary folder, and perform the following:

- Change the working directory to the temporary folder.
- Select File > Options from the menu, or click the toolbar button.

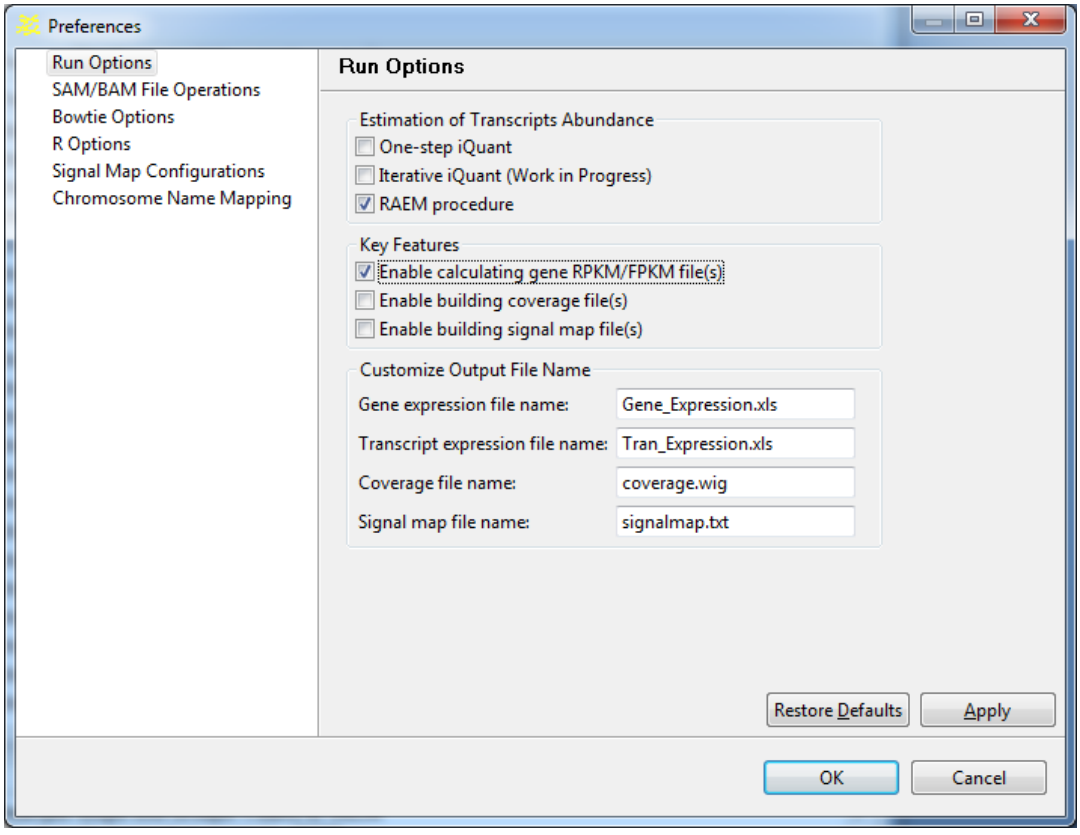
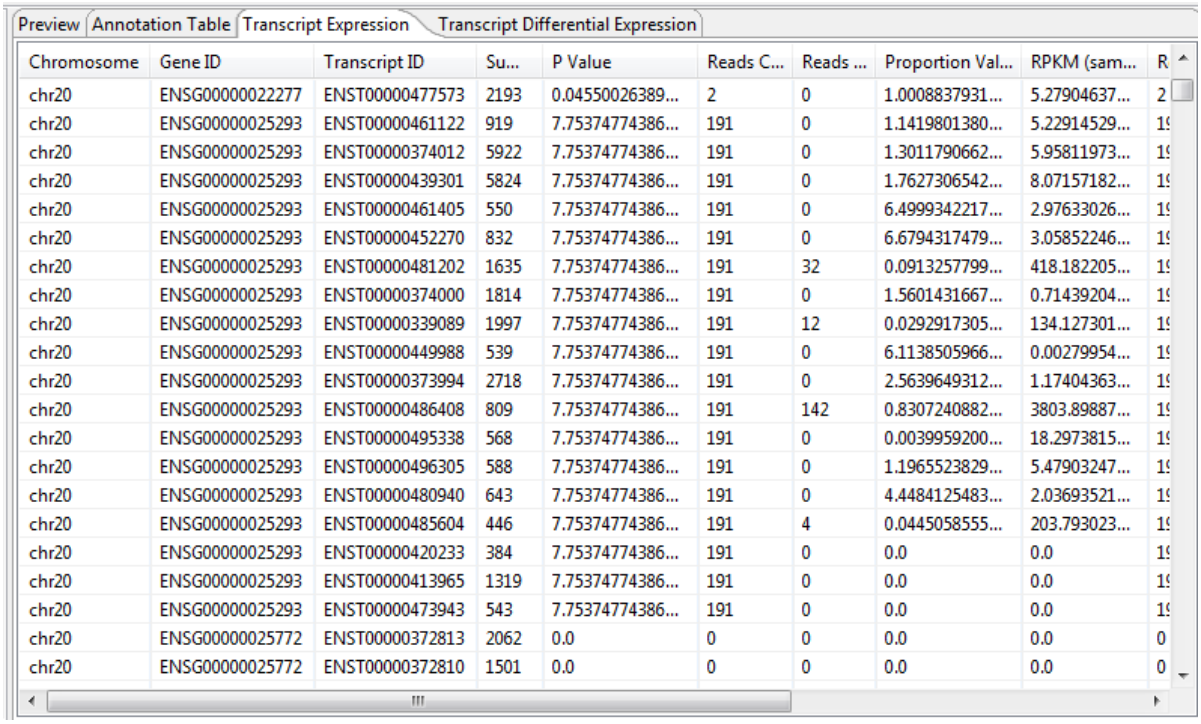


Figure 4.1.1-1 Options dialogue

- You can check the desired method to estimate transcripts abundance (See 3.2.1 Run Options).

- In the popup dialogue, you can customize the output file names.
- In the *File Browser Window*, right click on the desired annotation file and data file, and then left click *Add to Work Space*.
- In the *Work Space*, you can right click on the selected file, and then left click *Delete Selection* to remove selected file from *Work Space*.
- In the group A and group B tabs, you can assign the selected files into different groups if you want to export the P-value in the transcript expression file.
- Select Run > Run from the menu, or click the  toolbar button to calculate gene RPKM/FPKM file(s).
- After the process is finished, the annotation file, the genomic feature abundance scores and transcripts abundance scores are displayed in the *Navigator window*. Two Microsoft EXCEL compatible gene expression matrix and transcripts expression matrix are also generated in the temporary folder.



Preview | Annotation Table | Transcript Expression | Transcript Differential Expression

Chromosome	Gene ID	Transcript ID	Su...	P Value	Reads C...	Reads ...	Proportion Val...	RPKM (sam...	R
chr20	ENSG00000022277	ENST00000477573	2193	0.04550026389...	2	0	1.0008837931...	5.27904637...	2
chr20	ENSG00000025293	ENST00000461122	919	7.75374774386...	191	0	1.1419801380...	5.22914529...	19
chr20	ENSG00000025293	ENST00000374012	5922	7.75374774386...	191	0	1.3011790662...	5.95811973...	19
chr20	ENSG00000025293	ENST00000439301	5824	7.75374774386...	191	0	1.7627306542...	8.07157182...	19
chr20	ENSG00000025293	ENST00000461405	550	7.75374774386...	191	0	6.4999342217...	2.97633026...	19
chr20	ENSG00000025293	ENST00000452270	832	7.75374774386...	191	0	6.6794317479...	3.05852246...	19
chr20	ENSG00000025293	ENST00000481202	1635	7.75374774386...	191	32	0.0913257799...	418.182205...	19
chr20	ENSG00000025293	ENST00000374000	1814	7.75374774386...	191	0	1.5601431667...	0.71439204...	19
chr20	ENSG00000025293	ENST00000339089	1997	7.75374774386...	191	12	0.0292917305...	134.127301...	19
chr20	ENSG00000025293	ENST00000449988	539	7.75374774386...	191	0	6.1138505966...	0.00279954...	19
chr20	ENSG00000025293	ENST00000373994	2718	7.75374774386...	191	0	2.5639649312...	1.17404363...	19
chr20	ENSG00000025293	ENST00000486408	809	7.75374774386...	191	142	0.8307240882...	3803.89887...	19
chr20	ENSG00000025293	ENST00000495338	568	7.75374774386...	191	0	0.0039959200...	18.2973815...	19
chr20	ENSG00000025293	ENST00000496305	588	7.75374774386...	191	0	1.1965523829...	5.47903247...	19
chr20	ENSG00000025293	ENST00000480940	643	7.75374774386...	191	0	4.4484125483...	2.03693521...	19
chr20	ENSG00000025293	ENST00000485604	446	7.75374774386...	191	4	0.0445058555...	203.793023...	19
chr20	ENSG00000025293	ENST00000420233	384	7.75374774386...	191	0	0.0	0.0	19
chr20	ENSG00000025293	ENST00000413965	1319	7.75374774386...	191	0	0.0	0.0	19
chr20	ENSG00000025293	ENST00000473943	543	7.75374774386...	191	0	0.0	0.0	19
chr20	ENSG00000025772	ENST00000372813	2062	0.0	0	0	0.0	0.0	0
chr20	ENSG00000025772	ENST00000372810	1501	0.0	0	0	0.0	0.0	0

Figure 4.1.1-2 Navigator window

4.2. Calculating Genomic Feature Abundance Scores

SAMMate is compatible with both single-end and paired-end short reads mapped to exons (e.g. available in SAM/BAM format) to accurately estimate gene expression scores. *SAMMate* can also process RNA-seq data file in BED format.

SAMMate allows users to calculate the genomic feature abundance scores for any user-defined genomic intervals. This utility dramatically simplifies the technical barriers for discovering novel genes.

4.2.1 Example

Download the genome annotation file and the simulation RNA-seq data to a local hard drive from http://SAMMate.sourceforge.net/download.html. Unzip to a temporary folder, and perform the following:

- Change the working directory to the temporary folder.
- Select File > Options from the menu, or click the toolbar button.

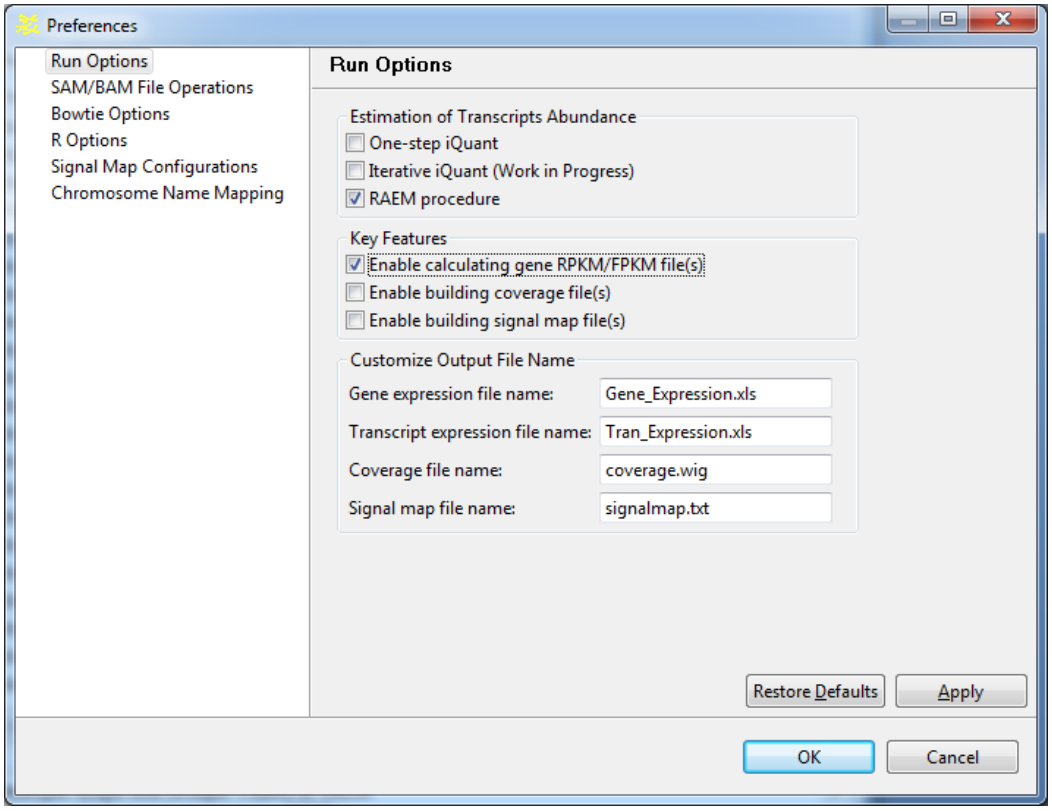
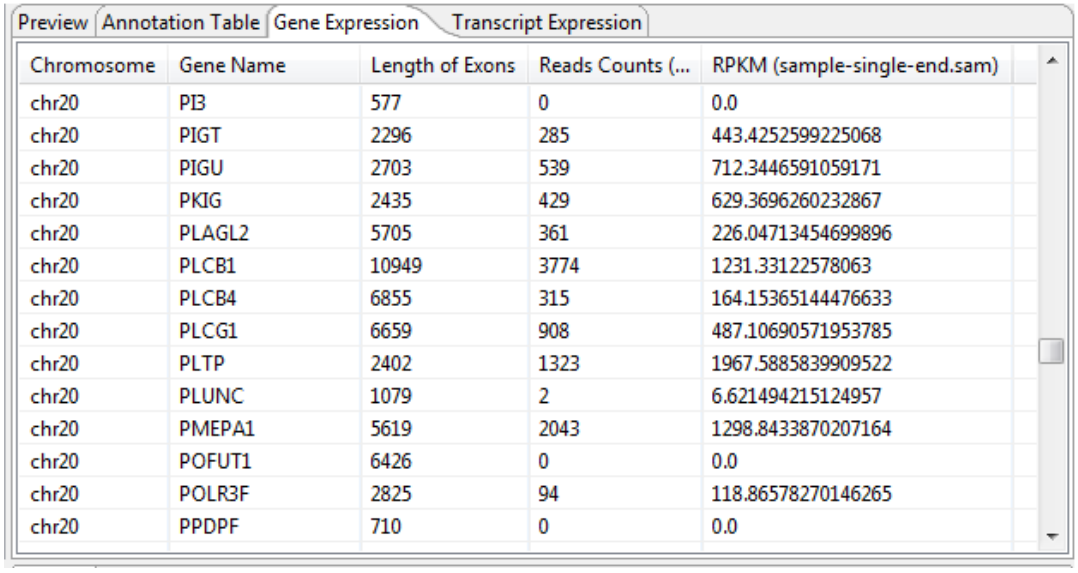


Figure 4.2.1-1 Options dialogue

- You can check the desired features to *enable calculating gene RPKM/FPKM file(s)* (See 3.2.1 Run Options).
- In the popup dialogue, you can customize the output file names.
- In the *File Browser Window*, right click on the desired annotation file and data file, and then left click *Add to Work Space*.
- In the *Work Space*, you can right click on the selected file, and then left click *Delete Selection* to remove selected file from *Work Space*.
- Select Run > Run from the menu, or click the toolbar button to calculate gene RPKM/FPKM file(s).
- After the process is finished, the annotation file and the genomic feature abundance scores are displayed in the *Navigator window*. A Microsoft EXCEL compatible gene expression matrix is also generated in the temporary folder.



Chromosome	Gene Name	Length of Exons	Reads Counts (...	RPKM (sample-single-end.sam)
chr20	PI3	577	0	0.0
chr20	PIGT	2296	285	443.4252599225068
chr20	PIGU	2703	539	712.3446591059171
chr20	PKIG	2435	429	629.3696260232867
chr20	PLAGL2	5705	361	226.04713454699896
chr20	PLCB1	10949	3774	1231.33122578063
chr20	PLCB4	6855	315	164.15365144476633
chr20	PLCG1	6659	908	487.10690571953785
chr20	PLTP	2402	1323	1967.5885839909522
chr20	PLUNC	1079	2	6.621494215124957
chr20	PMEPA1	5619	2043	1298.8433870207164
chr20	POFUT1	6426	0	0.0
chr20	POLR3F	2825	94	118.86578270146265
chr20	PPDPF	710	0	0.0

Figure 4.2.1-2 Navigator window

4.3. Generating a Signal Map for Peak Detection

A signal map is also another frequently demanded data format for NGS data analysis. In a signal map file, alignment results are represented in the per-base “pileup" format. In this format the single nucleotide short read coverage depth is calculated whereas the whole genome coverage is provided as a vector of integers with length 3.2×10^9.

SAMMate allows users to generate a signal map for a number of frequently performed sequential analyses to detect a wide range of genomic features based on user-defined genomic intervals.

4.3.1 Example

Download the genome annotation file and the simulation RNA-seq data to a local hard drive from http://SAMMate.sourceforge.net/download.html. Unzip to a temporary folder, and perform the following:

- Change the working directory to the temporary folder.
- Select File > Options from the menu, or click the toolbar button.
- Clicking *Signal Map Configurations*, users can customize the signal map intervals by adding the chromosome name, start position and end position. See Figure 4.3.1-1.

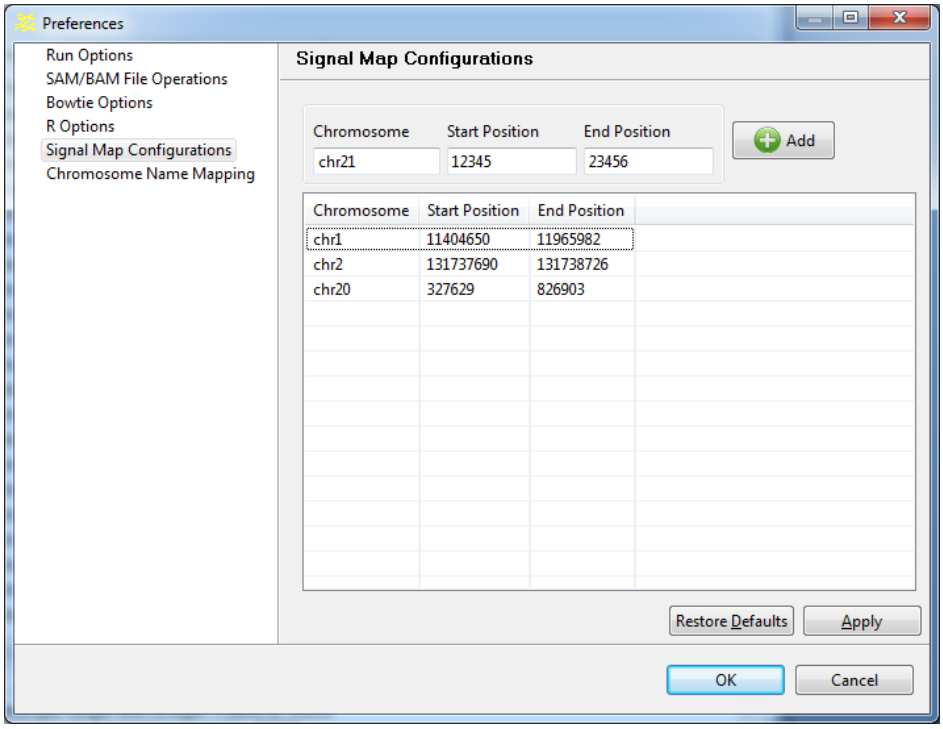


Figure 4.3.1-1 Options dialogue - Signal map configurations

- You can check the desired features to *building signal map file(s)* (See 3.2.1 Run Options).
- In the popup dialogue, you can customize the output signal file names.
- In the *File Browser Window*, right click on the desired annotation file and data file, and then left click *Add to Work Space*.
- In the *Work Space*, you can right click on the selected file, and then left click *Delete Selection* to remove selected file from *Work Space*.
- Select Run > Run from the menu, or click the  toolbar button to build signal map file(s).
- After the process is finished, the signal map file ordered by chromosome names is generated in the temporary folder.

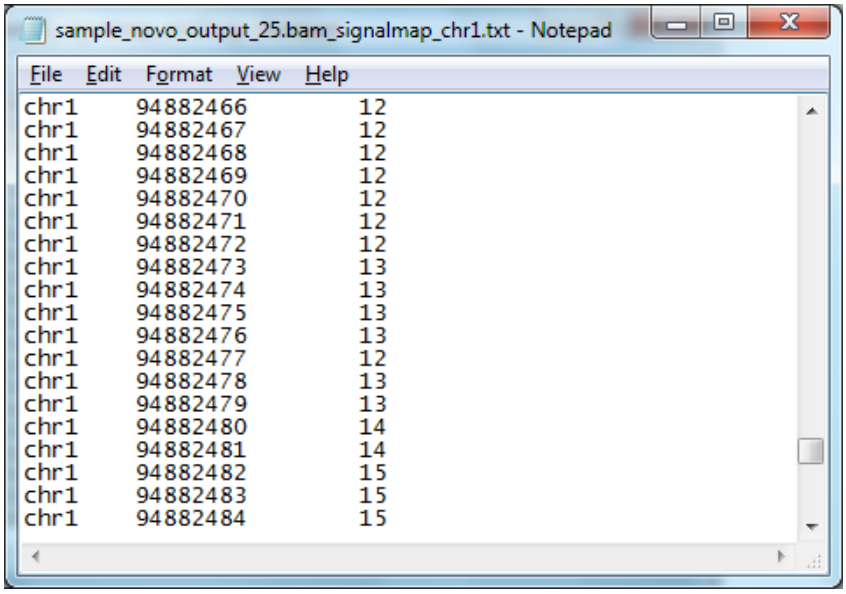


Figure 4.3.1-2 Signal map resulting file

4.4. Generating Wiggle File for Visualization

The wiggle (WIG) format is to display dense, continuous data such as GC percent, probability scores, and transcriptome data. A key *SAMMate* feature is to generate wiggle files for biomedical researchers so that they may visually search for gene structure alterations. These output files are compatible with the UCSC genome browser and other browsers used for visualization. This feature will allow biomedical researchers to visually check the alignment quality of selected genes in the selected genome regions.

4.4.1 Example

Download the genome annotation file and the simulation RNA-seq data to a local hard drive from http://SAMMate.sourceforge.net/download.html. Unzip to a temporary folder, and perform the following:

- Change the working directory to the temporary folder.
- Select File > Options from the menu, or click the  toolbar button.
- You can check the desired features to *building coverage file(s)* (See 3.2.1 Run Options).
- In the popup dialogue, you can customize the output coverage file names.
- In the *File Browser Window*, right click on the desired annotation file and data file, and then left click *Add to Work Space*.
- In the *Work Space*, you can right click on the selected file, and then left click *Delete Selection* to remove selected file from *Work Space*.

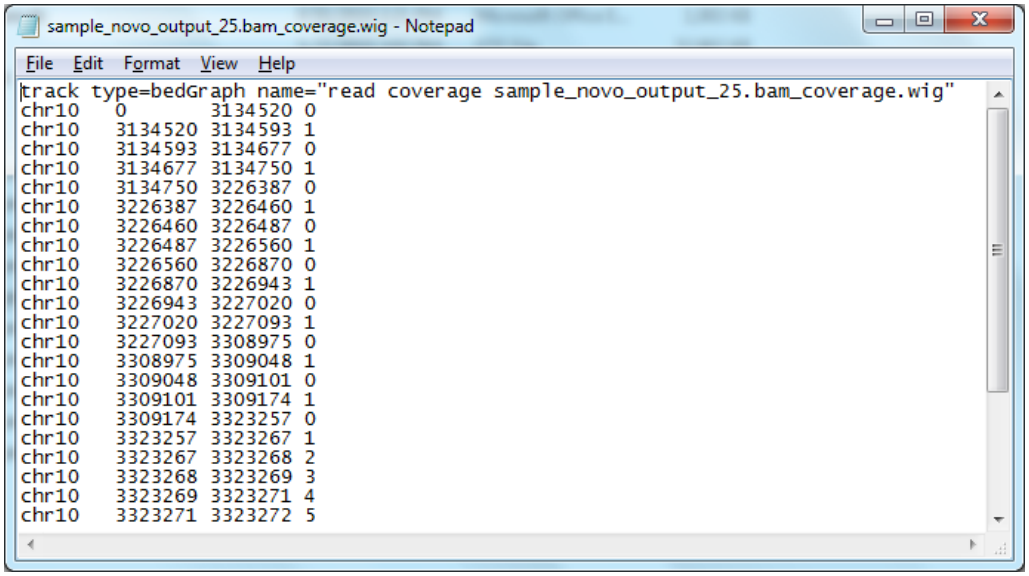


Figure 4.4.1-1 Coverage wiggle resulting file

- Select Run > Run from the menu, or click the  toolbar button to build coverage map file(s).
- After the process is finished, the coverage file sorted by chromosomal position is generated in the temporary folder.

4.5. Generating Alignment Report

Short read alignment statistics provide indispensable resources to examine the alignment quality as well as to compare the alignment results. *SAMMate* calculates and exports a number of alignment statistics including the percentage of uniquely mapped short reads as well as the percentage of short reads mapped to intergenic, exonic and intronic regions.

4.5.1 Example

Download the genome annotation file and the simulation RNA-seq data to a local hard drive from http://SAMMate.sourceforge.net/download.html. Unzip to a temporary folder, and perform the following:

- Change the working directory to the temporary folder.
- Select File > Options from the menu, or click the  toolbar button.
- You can check the desired features to *enable calculating gene RPKM/FPKM file(s)* (See 3.2.1 Run Options).
- In the *File Browser Window*, right click on the desired annotation file and data file, and then left click *Add to Work Space*.
- In the *Work Space*, you can right click on the selected file, and then left click *Delete Selection* to remove selected file from *Work Space*.
- Select Run > Run from the menu, or click the  toolbar.
- After the process is finished, a report file in HTML format is generated in the temporary folder.

SAMMate 2.6.1 Result Report:

Name	sample-single-end-GroupA-1.sam	sample-single-end-GroupB-1.sam
Reads Number On Exon	46230 (85.24173%)	46195 (85.196045%)
Reads Number On Intron	4342 (8.006048%)	4360 (8.041017%)
Reads Number On Junction	0 (0.0%)	0 (0.0%)
Reads Number On Intergenic	3662 (6.752222%)	3667 (6.7629375%)
Total Reads Number	54234	54222

Figure 4.5.1-1 Alignment report file